指数分布・累積分布関数・分位数

樋口さぶろお https://hig3.net

龍谷大学 先端理工学部 数理・情報科学課程

確率統計 I L03(2024-04-29 Mon)

最終更新: Time-stamp: "2024-04-29 Mon 15:49 JST hig"

今日の目標

- 累積分布関数から確率と分位数が求められる
- 指数分布の意味と確率密度関数が説明できる
- Google Collab 上で scipy で確率変数の計算ができる

L02-Q1
L02-Q2
Quiz 解答: 連続型確率変数

1. k 次のモーメントは,

$$\mathrm{E}[X^k] = \int_{-\infty}^{+\infty} x^k \cdot f(x)\,\mathrm{d}x = \int_0^{1/2} x^k 8x\,\mathrm{d}x = \frac{2^3 \cdot 2^{-k-2}}{k+2}.$$

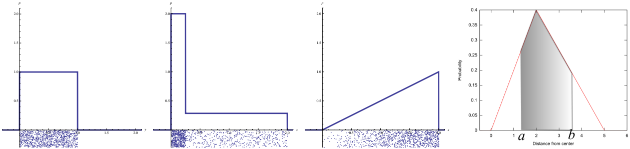
   $\mathrm{E}[X^0] = 1$ が確認できる.
2. $\mathrm{E}[X^1] = \frac{1}{3}$.
3. $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{1}{8} - (\frac{1}{3})^2 = \frac{1}{72}$.
4. $\mathrm{E}[(2X+3)^2] = 4\mathrm{E}[X^2] + 12\mathrm{E}[X^1] + 9\mathrm{E}[X^0] = 4 \times \frac{1}{8} + 12 \times \frac{1}{3} + 9 = \frac{27}{2}$.
5. $\mathrm{V}[2X+3] = 2^2\mathrm{V}[X] = \frac{1}{18}$.

ここまで来たよ

2 連続型確率変数

3 指数分布・累積分布関数・分位数
- 累積分布関数
- 指数分布
- 分位数
- 標本抽出と Python の scipy による確率変数の扱い

復習: 連続型確率変数の確率密度関数



横軸下の細かい点が, 標本 (縦方向の位置はランダムで意味なし)

命題 (確率密度関数と確率 岩薩林 確率・統計 (4.1)**)**

$$P(c < X \le d) = F(d) - F(c) = \int_c^d f(x)\,\mathrm{d}x \quad (\text{下側面積}) \tag{C3}$$

定義 (累積分布関数 岩薩林 確率・統計 (3.2)**)**

x の関数 $F(x) = P(X \le x)$ を確率変数 X の累積分布関数という.

$x \in \mathbb{Z}$ に限ると，数列 $F_x = F(x)$ の階差数列が $f_x = f(x)$.

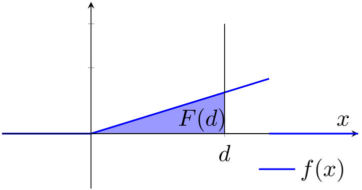
命題 (累積分布関数の性質 岩薩林 確率・統計 pp.50,51**)**

(C1) $F(x)$ は広義単調増加関数
(C2) $F(-\infty) = 0, F(+\infty) = 1$.
(C3) $P(c < X \le d) = F(d) - F(c)$.

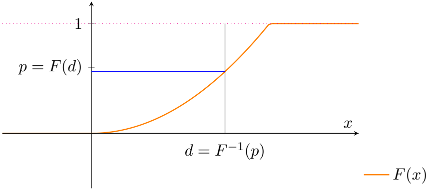
定義 (連続型確率変数の累積分布関数 岩薩林 確率・統計 (4.4))

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' \qquad \text{(C4 相当)}$$

確率密度関数 $f(x)$. 微分できる場合は $F'(x) = f(x)$.

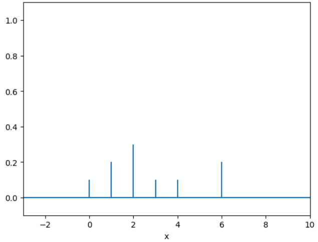


累積分布関数 $F(x)$

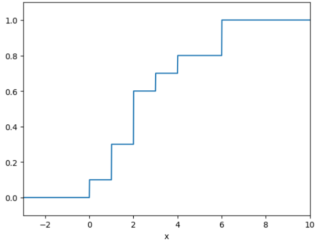


離散型確率変数の場合

確率関数 $p(x)$



累積分布関数は連続型と同様の定義 $F(x) = P(X \leq x)$



ここまで来たよ

2 連続型確率変数

3 指数分布・累積分布関数・分位数
- 累積分布関数
- 指数分布
- 分位数
- 標本抽出と Python の scipy による確率変数の扱い

指数分布 岩薩林 確率・統計 p.80

連続型確率変数の分布の, 名前がつく有名な例. パラメタ（母数）$\lambda > 0$ を持つ family (パラメタは確率変数でない. $8x^1$ の 8 や 1 のこと).

定義 (指数分布)

連続型確率変数 X が確率密度関数

$$f(x;\lambda) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

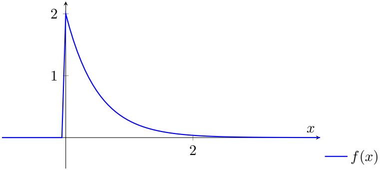
を持つとき, X はパラメタ $\lambda > 0$ の指数分布に従うといい, 記号 $\sim$(従う) で $X \sim \mathrm{Exp}(\lambda)$ と書く (Exp は科目ローカル記号)

性質 実は $\mathrm{E}[X^k] = \lambda^{-k}$ (多数回部分積分)
意味時間的にランダムに起きる事象，例えば,「ある機械が x 秒後に初めて故障する」「サッカーで x 秒後に次のゴールが起きる」のような事象の，時間間隔 x のしたがう分布.

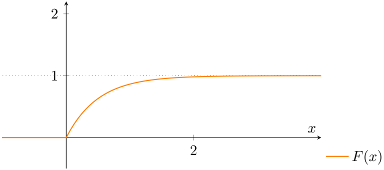
$X \sim \mathrm{Exp}(\lambda)$, 確率密度関数 $f(x) = \lambda \mathrm{e}^{-\lambda x}$ のとき，累積分布関数

$$F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' = \begin{cases} \int_{-\infty}^{x} 0\ \mathrm{d}x' = 0, & (x \leq 0) \\ \int_{-\infty}^{0} 0\ \mathrm{d}x' + \int_{0}^{x} \lambda \mathrm{e}^{-\lambda x'} \mathrm{d}x' = 1 - \mathrm{e}^{-\lambda x}. & (x > 0) \end{cases}$$

確率密度関数



累積分布関数



L03-Q1

Quiz(指数分布)

あるサッカーチームが，開始後 X 分で最初の得点をするとしたとき，この X はパラメタ $\lambda = 1/30$ の指数分布にしたがう（開始直後の特別なイベントは無視してゲームは時間的に一様とする．毎週の試合をつなげて考える, 本当は相手のチームによるだろうけど無視）.

1. 開始後 5 分から 10 分に最初の得点する確率を求めよう．有効数字 3 桁で答えよう.
2. 開始後 80 分たっても得点していない確率を求めよう．有効数字 3 桁で答えよう.
3. 得点と次の得点の間の時間の母平均値を求めよう．有効数字 3 桁で答えよう.
4. 1 点目がはいる確率が 2/3 を超える時刻を求めよう.

指数・対数関数の値は Windows の電卓．iPhone を横向きにして関数電卓で．または, Google Colab で np.exp(数値), np.log(数値).

ここまで来たよ

連続型確率変数の分位数関数

定義 (分位数関数 岩薩林 確率・統計 なし)

確率変数 X の 累積分布関数の, 区間 $[0,1]$ で定義された逆関数 $x = F^{-1}(y)$ を分位数関数, $F^{-1}(q)$ を, q-分位数 という.

意味

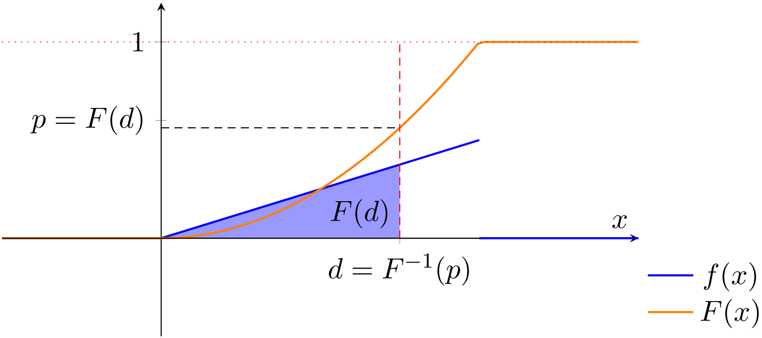
'q-分位数とは，確率 $P(X \leq d)$ が q に等しくなる**最小ぎりぎりの**境い目 $x = d$'

$F^{-1}(y)$ は, 確率 $y = q$ を与えると, そうなる境い目を返してくれる関数.

言い訳

定義域 $[0,1]$ の端は含められない場合もある.

$F(x)$ が狭義単調増加でない限り逆関数 F^{-1} は定義できない．逆像 $F^{-1}(y)$ は集合になるので $\inf F^{-1}(y)$ とするのが正確．ただし，$F^{-1}(0)$ は極限 $\lim_{\epsilon \to +0} F^{-1}(+\epsilon)$ とする.



例 (中央値・四分位数)

$F^{-1}(\frac{1}{2})$ は分布の (母) 中央値, $F^{-1}(\frac{1}{4}), F^{-1}(\frac{3}{4})$ は分布の (母) 四分位数.

データ分析

(標本) 中央値 岩薩林 確率・統計 p.27

ここまで来たよ

Python で連続型確率変数

- expon: 指数分布 (固有名詞)
- loc(ation): 分布の (平均と限らないけど) 位置を表す．指数分布は 0.
- scale: 分布の (標準偏差と限らないけど) 幅を表す．指数分布は $1/\lambda$.

指数分布

```
import scipy.stats
rvx = stats.expon(loc=0,scale=2) # Exp(1/2) loc=0, scale=1/λ
rvx.cdf(x) # cumulative density function 累積分布関数 F(x)
rvx.ppf(q) # percent point function 分位数関数(の類義語) F^{-1}(x

#
rvx.moment(4) # 4次のモーメント E[X^4]
rvx.expect(lambda x: x**2 + 3* x +1) # 母期待値 E[X^2 + 3X + 1]
```

Google Colab 連続型確率変数-指数分布.ipynb LearnMoodle

L03-Q2

Quiz(指数分布)

連続型確率変数 X_1, X_2 はそれぞれ確率密度関数

$$f_1(x) = \begin{cases} 2\mathrm{e}^{-2x} & (x > 0) \\ 0 & (x \le 0) \end{cases}, f_2(x) = \begin{cases} 2\mathrm{e}^{-2(x-3)} & ((x-3) > 0) \\ 0 & ((x-3) \le 0) \end{cases}$$

を持つ.

1. 確率密度関数のグラフを（重ねて）描こう.
2. $\mathrm{E}[X_1], \mathrm{E}[X_2]$ を求めよう.

関数で x を $x-3$ に置き換えると，グラフは横 (x) 方向に $+3$ 平行移動される

L03-Q3

Quiz(指数分布)

連続型確率変数 X_1, X_2 はそれぞれ確率密度関数

$$f_1(x) = \begin{cases} 2e^{-2x} & (x > 0) \\ 0 & (x \leq 0) \end{cases}, f_2(x) = \begin{cases} \frac{1}{3} \cdot 2e^{-2(x/3)} & ((x/3) > 0) \\ 0 & ((x/3) \leq 0) \end{cases}$$

を持つ.

1. 確率密度関数のグラフを（重ねて）描こう.
2. $\mathrm{E}[X_1], \mathrm{E}[X_2]$ を求めよう.

関数で x を $x/3$ に置き換えると，グラフは横 (x) 方向に 3 倍に拡大される